Task Demand Variation in Air Traffic Control: Implications for Workload, Fatigue, and Performance

Tamsyn Edwards[1] Cynthia Gabets[1], Joey Mercer[2], Nancy Bienert[1]

[1] San Jose State University/NASA Ames Research Center, Moffett Field, CA 94035, USA
[2] NASA Ames Research Center, Moffett Field, CA 94035, USA
{tamsyn.e.edwards, cynthia.wolter, joey.mercer, nancy.bienert}@nasa.gov

Abstract. In air traffic control, task demand and workload have important implications for the safety and efficiency of air traffic, and remain dominant considerations. Within air traffic control, task demand is dynamic. However, research on demand transitions and subsequent controller perception and performance is limited. This research uses an air traffic control simulation to investigate the effect of task demand transitions, and the direction of those transitions, on workload and fatigue and one efficiency performance measure. Results indicate that a change in task demand appears to affect both workload and fatigue ratings, although not necessarily performance. In addition, participants' workload and fatigue ratings in equivalent task demand periods appear to change depending on the demand period preceding the time of the current ratings. Further research is needed to enhance understanding of demand transition and workload history effects on operator experience and performance, in both air traffic control and other safety-critical domains.

Keywords: Workload transitions · Workload history · Air traffic control · Fatigue · Time based metering · Demand transitions

Introduction

Within the safety critical domain of air traffic control (ATC), workload "is still considered one of the most important single factors influencing operators' performance" [1 p639]. Workload has been defined within the ATC domain as the result of an interaction between task demand and the controllers' selected strategy [2]. The association of workload and controller performance has important implications for the safety and efficiency of air traffic (e.g. [3]; [4]). Workload therefore remains a dominant consideration.

In ATC, as with many other safety critical environments, task demand and workload are dynamic. Air traffic controllers (ATCOs) frequently experience changes in traffic counts and complexity, potentially resulting in experienced transitions between high and low workload. These transitions can be expected by the controller, such as when traffic counts change based on time of day or known activities in surrounding sectors, or unexpected, for example, through increased complexity resulting from an emergency situation. Transitions may also be gradual or sudden [5].

Research on workload transitions and performance is limited however, with workload studies frequently utilizing a constant task demand or workload [6]. Of the research available, there appears to be conflicting findings. [7] reported that overall performance efficiency on a vigilance task was not affected by task demand transitions, whether the transition was expected or unexpected. However, others have found performance on vigilance tasks to be influenced by a low to high demand transition (e.g. [8]) or high to low demand transition (e.g. [5]). Task demand and workload transition research specific to an ATC environment is particularly underrepresented, potentially limiting the application of findings. Consequently, there is limited understanding of the influence of workload transitions on performance in an air traffic environment.

The aim of this study was therefore to investigate the influence of expected and gradual task demand transitions (high-low-high and low-high-low) on workload, fatigue and performance, within a high fidelity ATC simulation environment. Due to the quantity of measures, data and analyses generated from this study, only a subset of the measures and findings that are most relevant to this research aim are presented.

Method

Design

An en-route air traffic control (ATC) human in the loop (HITL) simulation was utilized to investigate task demand variation and workload, fatigue, and performance. Efficiency-related performance was inferred from seconds of aircraft delay. Participants were eight ex-ATCOs who had worked in enroute airspace in Oakland Air Route Traffic Control Centre (ARTCC). Pseudo pilots were paired with controllers. Participants operated a combined low and high altitude sector, and were assigned to meter aircraft into the northeast corner of the Phoenix Terminal Radar Approach Control (TRACON). This airspace was selected for the complex mix of arrivals and overflights in the airspace.

The study used a within measures design. The direction of the task demand transition was manipulated to create two scenarios. Scenario 1 followed a high-low-high task demand pattern and scenario two followed a low-high-low task demand pattern. The creation of three task demand periods was implemented in order to better reflect the multiple task demand transitions that can be experienced within an operational environment. In addition, this permitted an extension of previous studies that had focused on the comparison of one transition period (e.g. [5]). Each simulation session lasted for 90 minutes and consisted of three, 20 minute [9] periods of alternating task demand interspersed with a total of three, 10 minute transition phases. Task demand was created by changing the number of aircraft under control [10] and the complexity of the task by the number of arrival aircraft as opposed to overflights. Task demand phases in the high-low-high and low-high-low scenarios were equivalent, permitting comparability between demand variation scenarios. Scenarios followed a counterbalanced presentation. Participants were required to complete all control actions and meter aircraft to arrive at a scheduled time. Participants were provided with a 1 hour briefing prior to the start of the study, and six training rungs (four 90 minute training runs and two 45 minute training runs).

Participants

A total of eight male participants took part in the simulation. Age ranged from 50 years – 64 years. Participants responded to grouped age ranges and so an average age could not be calculated. All participants were ex-controllers. Participants had worked as en-route controllers in the Oakland ARTCC. Years of experience as an ATCO (excluding training) ranged from 22 – 31 years (M=26.56, SD=3.90).

Measures and Apparatus

Covariate factors were measured using subjective, self-report scales. Mental workload was measured using the uni-dimensional Instantaneous Self-Assessment scale (ISA) [11]. Every 3 minutes, participants were presented with the ISA rating scale at the top of the radar scope and asked to click on the workload rating. Fatigue was measured using the Samn-Perelli scale [12]. Fatigue measures were taken 3 minutes into each 20 minute task demand phase, three minutes prior to the end of each 20 minute task demand phase, and six minutes into each 10 minute transition phase. This periodicity was selected to capture data across each stable task demand period, and refined based on results from three pilot studies. Performance was assessed by aircraft delay at three nautical miles from the meter fix point.

The software used was the Multi-Aircraft Control System (MACS). Participant workstations were configured with a BARCO large-format display and keyboard/trackball combination that emulates what is currently used in en-route air traffic control facilities. Voice communications were enabled via a custom, stand-alone system. Datalink communications were also available. Data were collected continuously through MACS's data collection processes.

Results

Analysis approach and reporting

Due to the quantity of analyses and findings, only the most relevant data trends are presented in this paper. Comparisons of only the three 20 minute task demand periods per scenario are presented below. Comparison between task demand periods was a priority to address the research aim. In addition, one dependent variable – aircraft arrival delay - was assessed as a measure of participants' efficiency-related performance. An efficiency-related performance measure was selected for analysis as opposed to a safety-related performance measure as previous research suggests that controllers can maintain safety-related performance without observed significant changes even under high periods of demand by applying workload management strategies. Changes in performance are therefore frequently first observed in efficiency-related tasks (e.g. [13]). An efficiency-related task was therefore potentially more sensitive to changes in performance.

For each covariate factor and dependent variable, descriptive statistics were first reviewed, followed by further exploration through the application of two repeated measures ANOVAs – one for each task demand transition scenario (scenario 1: high-low-high demand; scenario 2: low-high-low demand). The decision to apply separate repeated measures one-way ANVOAs was made based on the research aims and a review of previous research analysis approaches to similar experimental designs (e.g. [5]). The research aim focused on investigating the effect of task demand on covariate and performance variables, including the direction of the task demand. One way ANOVAs permitted exploration of changes within each task demand scenario. Prior to all inferential statistics, data were checked for normality and sphericity violations. Unless otherwise reported, all data met these assumptions.

Task Demand Variation Manipulation Check

A review of the descriptive statistics suggests that task demand did vary in the direction intended (Fig.1). Figure 1 confirms that the number of aircraft in the controller sector, and number of arriving aircraft (creating complexity) were similar between equivalent task demand periods regardless of scenario (high-low-high demand or low-high-low demand).

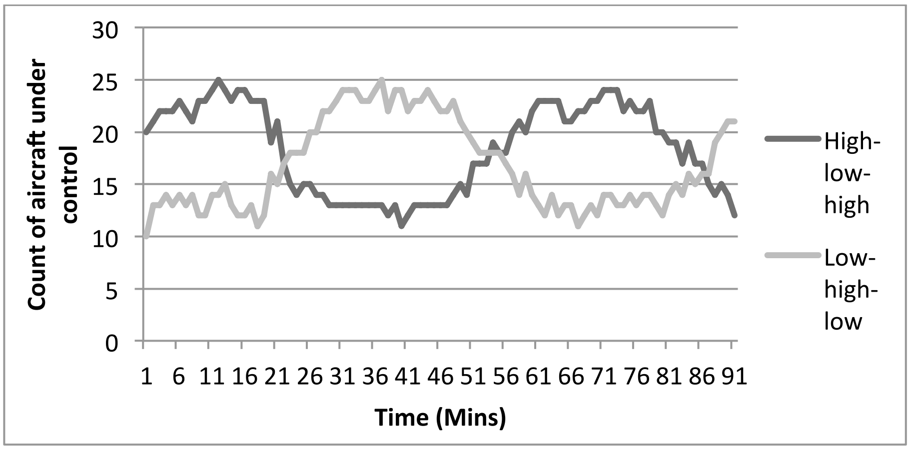

Fig. 1. Count of aircraft under control by minute for scenario 1 (*high-low-high demand*) and scenario 2 (low-*high-low demand*).

Task Demand and Subjectively Experienced Factors

Task Demand and Workload

Table 1. Mean and standard deviation for workload (as rated by ISA) in both scenario 1 and scenario 2, averaged across 20 minute task demand periods.

Workload (ISA)	Task demand period 1 (0-20 minutes)		Task demand period 2 (31-50 minutes)		Task demand period 3 (61-80 minutes)	
	M	SD	M	SD	M	SD
Scenario 1 workload (High-low-high)	3.67	0.77	2.87	0.61	3.85	0.62
Scenario 2 workload (Low-high-low)	2.78	0.64	4.06	0.71	3.33	0.61

Workload ratings were averaged across the stable task demand 20 minute periods for analysis to facilitate comparison between the separate task demand periods. A review of the descriptive statistics (Table 1) suggest that workload in both scenarios varied as expected with task demand. In scenario 1 (high-low-high demand) workload appears to be rated slightly higher in the third task demand period (high demand) (M=3.85, SD=0.62) compared to the first task demand period (high demand) (M=3.67, SD=0.77). In scenario 2 (low-high-low demand), workload was rated highest in the high demand, second task demand phase (M=4.06, SD=0.71). However, on average, participants rated perceived workload to increase in the third task demand period (low demand) (M=3.33, SD=0.61) compared to the first low demand period (M=2.78, SD=0.64). Comparing between scenario 1 and 2, the high demand period is perceived to generate the most workload for participants in the low-high-low demand scenario (M=4.06, SD=0.71), although the high demand periods were objectively equivalent between scenarios. Comparing across low demand periods between conditions, workload is rated similarly in the first period of scenario 2 (M=2.78, SD=0.64) and the middle period of scenario 1 (M=2.87, 0.61). However, the low demand period in the third period of scenario 2 is rated as higher workload than either of the other low demand periods (M=3.33, SD=0.61).

To further examine the changes in perceived workload, a one-way repeated measures analysis of variance (ANOVA) was conducted for each scenario [5]. A one-way ANOVA was applied to each scenario, rather than one two-way ANOVA, as a direct statistical comparison of the interaction of the conditions would not provide the granularity of results required. To understand differences within-scenarios, a one way ANOVA was the most appropriate analysis. In relation to scenario 1 (high-low-high demand a significant main effect of task demand period was found on self-reported workload $F(2,14) = 44.23$, $p<0.001$. Pairwise comparisons revealed that workload was significantly lower in task demand period 2 (low demand) (M=2.87, SD=0.61) than high task demand period one (M=3.67, SD=0.77) ($p<0.005$) and three (M=3.85, SD=0.62) ($p<0.001$). Workload was not rated significantly differently between high demand period 1 (M=3.67, SD=0.77) and high demand period 3 (M=3.85, SD=0.62) ($p=0.68$). In scenario 2 (low-high-low demand) a significant main effect of task demand period was found on self-reported workload $F(2,14) = 32.72$, $p<0.001$. Pairwise comparisons revealed that workload was rated significantly higher in the high demand period (M=4.06, SD=0.71) than the first low demand period (M=2.78, SD=0.64) ($p<0.001$) and second low demand period (M=3.33, SD=0.61) ($p<0.005$). It was also identified that the workload ratings in the second low demand period (M=3.33, SD=0.61) were significantly higher than the first low demand period (M=2.78, SD=0.64), ($p<0.05$).

Task Demand and Fatigue

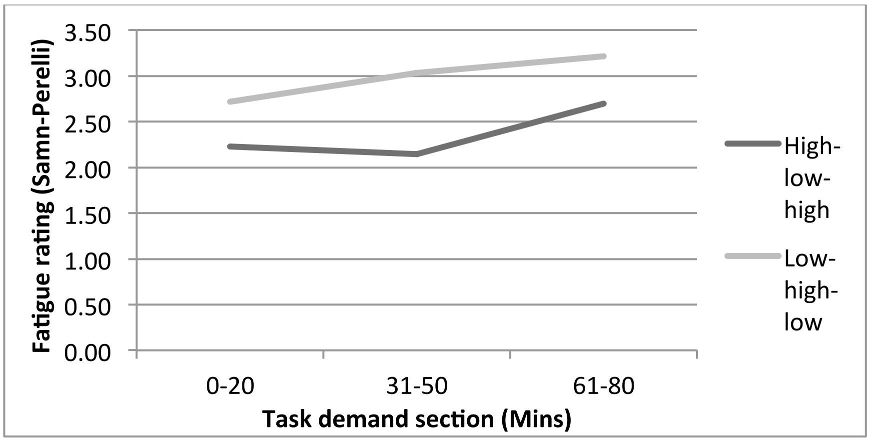

Fig. 2. Fatigue ratings (as measured by Samn-Perelli scale) averaged across 20 minute task demand periods for scenario 1 (*high-low-high demand*) and scenario 2 (*low-high-low demand*).

A review of the means of reported fatigue for each task demand period in scenario 1 (high-low-high demand) revealed that ratings of fatigue appeared similar between high demand period one (M=2.23, SD=0.71) and low demand period one (M=2.15, SD=0.77) (Fig. 2). Fatigue ratings were slightly higher in the third demand period, high demand period two (M=2.70, SD= 1.08). Conversely, in scenario 2 (low-high-low demand) fatigue ratings appeared to increase across each task demand period (first low task demand period: M=2.71, SD=1.01; first high task demand period: M=3.03, SD=1.42; second low task demand period: M=3.22, SD=1.54) (Fig. 2).

A one way ANOVA was utilized to explore the effect of task demand on fatigue ratings for both scenarios. In scenario 1 (high-low-high demand) Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2(2) = 9.44$, $p<0.01$. When considering this main effect, therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (E=0.56). No significant differences between fatigue ratings were identified $F(1.12, 7.81) = 2.48$, $p>0.05$. Differences between fatigue ratings in scenario 2 (low-high-low demand) approached significance, $F(2,14) = 3.40$, $p<0.1$. A further review of the descriptive data revealed that averaging across the two fatigue measures per task demand period (one three minutes into the period, and one three minutes before the end of the period) may be masking effect of task demand on fatigue. Participants' fatigue rating was frequently lower for the first measurement compared to the second measurement of the task demand period. Therefore, ANOVAs were repeated on two fatigue measurements per workload period. In scenario 1 (high-low-high demand) Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2(14) = 26.82$, $p<0.05$. When considering this main effect, therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (E=0.44). No significant differences between fatigue ratings were identified $F(2.18, 15.22) = 2.82$, $p>0.05$. The ANOVA applied to scenario 2 revealed a main effect of task demand on fatigue ratings $F(5,35) = 2.69$, $p<0.05$. Pairwise comparisons revealed that fatigue ratings were significantly lower at the first fatigue measurement of the first low task demand period (M=2.63, SD=1.06) compared to fatigue ratings in the second low task demand period (first fatigue measurement M=3.13, SD= 1.46, p=0.05; second fatigue measurement M=3.31, SD=1.65), $p<0.05$). No other differences were significant.

Task demand and performance

Table 2. Mean and standard deviation for arrival aircraft delay (in seconds) in both scenario 1 and scenario 2, averaged across 20 minute task demand periods.

Arrival aircraft delay (secs)	Task demand period 1 (0-20 minutes)		Task demand period 2 (31-50 minutes)		Task demand period 3 (61-80 minutes)	
	M	SD	M	SD	M	SD
Scenario 1 Aircraft delay (High-low-high)	13.88	5.32	7.70	3.6	-1.71	6.92
Scenario 2 Aircraft delay (Low-high-low)	10.48	3.07	9.93	2.54	7.50	4.86

A review of the average delay across 20 minute task demand periods in scenario 1 (high-low-high demand) (Table 2) suggests that participants reduced average aircraft delay across the task demand periods until aircraft were arriving early in the final task demand period (Table 2). The same pattern was seen in scenario 2 (low-high-low demand), although smaller reductions in delay are observed (Table 2). However, in both scenarios, performance variability appears to increase in the final task demand period, indicated by comparatively large standard deviations (Table 2). Data in scenario 1 (high-low-high demand) were further examined with a repeated measures ANOVA. A significant main effect of task demand period was found on arrival delay $F(2,14) = 12.84$, $p<0.005$. Pairwise comparisons revealed that aircraft delay was significantly longer in the first high demand period (M=13.88, SD=5.32) than the first low demand period (M=7.70, SD=3.6) ($p<0.05$) and the second high demand period (M=-1.71, SD=6.92) ($p<0.01$). Delay was also significantly longer in the first low demand period (M=7.70, SD=3.6) than the second high demand period (M=-1.71, SD=6.92) ($p<0.05$). Data in scenario 2 (low-high-low demand) were also further examined with a repeated measures ANOVA. No significant differences in arrival aircraft delay were identified $F(2,14) = 3.04$, $p>0.05$.

Discussion

Task demand varied as intended. Descriptive statistics confirmed that equivalent demand periods, regardless of scenario or position, were composed very similarly in terms of controlled aircraft count and arrival aircraft count. This suggests that changes in the covariates or dependent variable are unlikely to be attributed to demand differences between the created scenarios.

As expected, a change in task demand appears to affect both workload and fatigue ratings. Significantly different workload and fatigue ratings were reported within scenario, across task demand periods. However, a key finding of interest is that perception of workload and fatigue appear to differentiate depending on the demand period preceding the current ratings, in line with previous findings [5]. This finding is observed in the average workload ratings for each task demand period within scenarios (Table 1). In the first scenario (high-low-high task demand), workload is not perceived significantly differently between the first and second high task demand periods. Workload is rated as significantly lower during the low demand period compared to the high demand periods, however. Comparatively, in scenario 2 (low-high-low demand) workload is perceived to be significantly greater in the second low demand period than the first, potentially suggesting that workload is perceived to be greater after the high demand period. In addition, it is interesting to note that workload was perceived to be higher in the high demand period of scenario 2 than either of the high task demand periods in scenario 1.

This collection of findings indicates that the workload appears to be perceived differently depending on what precedes the time of rating. More specifically, results suggest that in this ATC task, a demand transition pattern of low-high-low demand may result in operators perceiving subsequent high and low demand periods after the initial low demand period as generating a greater workload than equivalent demand periods in a high-low-high demand transition pattern. A similar pattern of findings was seen in participants' fatigue ratings. In scenario 1 (high-low-high demand), fatigue ratings were not significantly different between demand phases. Fatigue ratings did increase in the final high demand period, although not significantly. In contrast, in scenario 2, participants reported on average that fatigued increase with each subsequent task demand period.

Although there is a lack of common agreement regarding the mechanisms by which task demand transitions may impact covariate factors [14], this collection of workload and fatigue findings may be interpreted in the context of Limited Resource theory [15] and arousal theories. Potentially, in scenario 1, the low demand period may have enabled controllers to use this time to recover resources and prepare for the next high task demand period. [16] has previously documented that this is an active control strategy controllers- use during low demand periods, when it is considered safe to do so. This recovery period may then limit the increase of perceived fatigue in the final high task demand period. Arousal theories may provide some insight into why this effect may not be seen in the low-high-low demand transition pattern. Arousal theories suggest that low workload (or underload) may lead to lower arousal, which may limit attentional resources and create boredom and lack of motivation. If a human operator started a task from this point, it may be that the following demand periods are perceived to be more demanding or fatiguing. By the final low demand period, the operator may find it difficult to pay attention. Attentional resources theories suggest however that if preceded by a higher demand, lower demand periods can be utilized to replenish attentional resources, not necessarily reducing arousal to a level that would create negative effects. The application of these theories therefore potentially account for the disparate findings between the different task demand transition patterns.

Performance did not appear to be negatively affected in relation to task demand variation, with delay times reducing across task demand periods within each scenario. Performance variability did increase however across task demand period, as inferred from increasing standard deviations. This pattern of findings for performance measures has also been documented previously, although for vigilance-based performance [7]. Controllers are not passive in their environment, and will utilize strategies to support performance [17]. The finding of improved aircraft arrival time may therefore be the result of controllers applying strategies to support performance across the demand periods.

Although not a direct focus of this research, it is important to note that this finding highlights an important issue for future research considerations. Although this measure of performance indicates that performance in terms of aircraft arrival time was maintained, and even improved, in scenario 2, controllers also reported greater perception of workload and fatigue. It is therefore possible that controllers may have experienced having to work harder to maintain performance, even though this was not observable in the performance measure itself. This result emphasizes that in order to detect, and prevent, performance declines, further research should focus on measures that are sensitive to the operators' experience, and that can be monitored and utilized to detect potential performance decline prior to a performance related incident.

It is acknowledged that these results are provisional, and results need to be interpreted within context. For example, in an air traffic environment, it is easier for the controller to build a picture of the traffic by ramping up with the traffic rather than just starting a session in a high demand period [16]. However, findings do have important implications for the prediction of controller performance in an operational environment. Findings suggest that high and low demand periods can affect controller perception of covariate factors such as workload and fatigue differentially depending on what has happened prior to the current situation. Supervisors may need to pay close attention to the number and direction of transitions that a controller experiences per session to most effectively support controller performance. Future research should further explore the relationship between previous task demands and the relationship on present controller experience, including the exploration of sudden, and unexpected, transitions. This enhanced understanding may have important implications for adaptive automation technologies that can support operator performance. In addition, a more detailed understand may facilitate more accurate predictions of performance in high and low demand periods, with important implications for identifying and preventing potential performance declines and associated performance-related incidents.

Conclusion

The effect of task demand transitions on covariate factors of workload and fatigue and one efficiency related performance measure was investigated within the context of an air traffic control task. Initial findings suggest that task demand variations affect participants' perceptions of workload and fatigue, although the effect of appears to be influenced by the direction of the previous demand periods. Performance appeared to be maintained across the control session. Previous research has infrequently considered transitions of task demand in an applied environment. Findings are consistent with the description of workload history effects [5], and that equivalent task demand periods can elicit different experiences for a human operator depending on what precedes the time of rating. Further research is required to enhance understanding of demand transition and history effects. Practical applications include guidance for operations room supervisors, and implications for predictions of performance in high and low demand periods, with important implications for identifying and preventing potential performance declines and associated performance-related incidents.

References

1. Di Nocera, F., Fabrozo, R., Terenzi, M., & Ferlazzo, F.: Procedural Errors in Air Traffic Control: Effects of Traffic Density, Expertise, and Automation. Aviat. Space Envir. Md., 77, 639--643 (2006)
2. Djokic, J., Lorenz, B., & Fricke, H.: Air Traffic Control Complexity as Workload Driver. Transport Res. C-Emer., 18, 930--936 (2010)
3. Redding, R.E.: Analysis of Operational Errors and Workload in Air Traffic Control. In Human Factors and Ergonomics Society Annual Meeting Proceedings, 36(17), 1321--1325. Human Factors and Ergonomics Society (1992)
4. Schroeder, D.J., Nye, L.G.: An Examination of the Workload Conditions Associated with Operational Errors/Deviations at Air Route Traffic Control Centers. Washington, DC: Federal Aviation Administration Office of Aerospace Medicine (1993)
5. Cox-Fuenzalida, L.E.: Effect of Workload History on Task Performance. Hum. Factors, 49, 277--291 (2007)
6. Hancock, P.A., Williams, G., Manning, C.M.: Influence of Task Demand Characteristics on Workload and Performance. Int. J. Aviat. Psychol., 5, 63--86 (1995)
7. Helton, W.S., Shaw, T., Warm, J. S., Matthews, G., Hancock, P.: Effects Of Warned and Unwarned Demand Transitions on Vigilance Performance and Stress. Anxiety Stress Copin., 21, 173--184 (2008)
8. Moroney, B.W., Warm, J.S., & Dember, W.N.: Effects of Demand Transitions on Vigilance Performance and Perceived Workload. In Proceedings of the Human Factors And Ergonomics Society Annual Meeting 39, 1375--1379. SAGE Publications (1995, October)
9. Galster, S.M., Duley, J.A., Masalonis, A.J., Parasuraman, R.: Air Traffic Controller Performance and Workload Under Mature Free Flight: Conflict Detection and Resolution of Aircraft Self-Separation. Int. J. Aviat. Psychol., 11, 71--93 (2001)
10. Tenney, Y.J., Spector, S.L.: Comparisons of HBR models with Human-in-the-Loop Performance in a Simplified Air Traffic Control Simulation with and without HLA Protocols: Task Simulation, Human Data and Results. In: Proceedings of the 10th Conference on Computer Generated Forces & Behaviour Representation, Norfolk, VA, (2001)
11. Tattersall, A.J., Foord, P.S.: An Experimental Evaluation of Instantaneous Self-Assessment as a Measure of Workload. Ergonomics, 39, 740–-748 (1996)
12. Samn, S.W., Perelli, L.P.: Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations (No. SAM-TR-82-21). School of Aerospace Medicine Brooks Afb Tx (1982)
13. Edwards, T. Human performance in air traffic control. PhD diss., University of Nottingham (2013)
14. Wickens, C.D., Mabor, A.S., & McGee, J.P.: Flight to the Future: Human Factors in Air Traffic Control. National Academy Press Washington D.C. (1997)

15.Wickens, C.D.: Engineering Psychology and Human Performance. Harper Collins, New York (1992)
16. Edwards, T., Sharples, S., Kirwan, B., Wilson, J.R.: Identifying Markers of Performance Decline in Air Traffic Controllers. In: Di Bucchianico, G., Vallicelli, A., Stanton, N.A., Landry, S.J. (eds.) Human Factors in Transportation: Social and Technological Evolution Across Maritime, Road, Rail, and Aviation Domains (In Press)
17. Sperandio, J.C.: Variation of Operator's Strategies and Regulating Effects on Workload. Ergonomics, 14, 571--577 (1971)